1 PROOFS

1.1 Theorem 2.4

We assume that there is a possible solution. Given that $\mathcal{T}_i$ and $\mathcal{S}_i$ are executable within polynomial time, we can also verify whether the solution is correct or not in polynomial time. Thus, the problem is in NP. To prove the NP-complete, we show a reduction from the decision problem to the set cover problem, one of Karp's 21 NP-Complete problems [?].

We construct an instance of repair problem as follows. Let $n = min(p, q)$, $m = 2$, $\eta = n$, $x_i = 0$ for $1 \le i \le n$. For trend filters $\mathcal{T}_i$ and seasonality filters $\mathcal{S}_i$, $t_{j+1} = \mathcal{T}_{j+1}(x_{j+1}, x_j) = x_{j+1} - x_j$ if $u_i \in Y_j$ and $(i-j) \nmid 2$ otherwise 0. $s_i = \mathcal{S}_i\left(\left[d_j;\ 1 \le j \le n \wedge (j-i) \mid m\right]\right) = -n - 1 + 2\sum_{u_i \in Y_j, (i-j)|2}(x_j - t_j) + 2\sum_{u_i \in Y_j, (i-j)\nmid 2}(x_{j+1} - t_{j+1})$ for $1 \le i \le n$. If $p > q$, we add some artificial x_i, t_i, s_i where $x_i = 0, t_i = \mathcal{T}_i(x_i, x_{i-1}) = 0, t_i = \mathcal{S}_i\left(\left[d_j;\ 1 \le j \le n \wedge (j-i) \mid 2\right]\right) = 0$ for $n < i \le p$, i.e., $r_i = 0 \le n$ is always satisfied. If $p < q$, we add some artificial x_i where $x_i = 0$ for $n < i \le q$. Finally, we set the repaired value of x_i, i.e., $x_i' = 1$. Figure 1 shows an example. Given $Y_3 = \{u_1, u_6\}$, $p = 6$, $q = 7$, we construct an instance with $n = min(p, q) = 6$, $x_i = 0$ for $1 \le i \le 7$. Since $u_6 \in Y_3$ and $(6-3) \nmid 2$, $t_4 = \mathcal{T}_4(x_3, x_4) = x_4 - x_3$. Otherwise $t_i = 0$ for $i \in \{1, 2, 3, 5, 6\}$. $s_1 = \mathcal{S}_1(x_1 - t_1, x_3 - t_3, x_5 - t_5) = -n - 1 + 2(x_3 - t_3) = -n - 1 + 2x_3$. $s_6 = \mathcal{S}_6(x_2 - t_2, x_4 - t_4, x_6 - t_6) = -n - 1 + 2(x_4 - t_4) = -n - 1 + 2x_3$. Otherwise $s_i = -n + 1$ for $i \in \{2, 3, 4, 5\}$. We can derive that $r_1 = x_1 - t_1 - s_1 = n + 1 + x_1 - 2x_3$, $r_6 = x_6 - t_6 - s_6 = n + 1 + x_6 - 2x_3$, $r_4 = n + 1 + x_3$, $r_1 = r_2 = r_5 = n + 1$ Therefore, when $x_3 = 1$ in the gray box, $r_1 = r_6 = n - 1 \le \eta$ in the blue box, while $r_4 = n + 2 > \eta$, $r_1 = r_2 = r_5 = n + 1 > \eta$. It is clear that we can construct $x_i, \eta, \mathcal{T}_i$, $\mathcal{S}_i$ in polynomial time.

We now show that $\Delta(x, x') \le \tau$ if and only if $card(L) \le \tau$. Begin by assuming $card(L) \le \tau$ has a solution L, repair x_i to $x_i' = 1$ for each $Y_i \in L$. Thereafter, for $u_j \in Y_i$, it is observed that $-n+1 \le s_j \le n-1$. If $u_i \in Y_{j-1}$ and $(i-j) \nmid 2$ $t_j = \mathcal{T}_j(x_j, x_{j-1}) = x_j - x_{j-1}$ We can derive that $r_j = x_j - t_j - s_j = x_j + s_j$ or $x_{j-1} + s_j$, then $-n+2 \le r_j \le n$. Since $\eta = n$, $\mid r_j \mid \le \eta$. In summary, $\mid r_j \mid \le \eta$ if $u_j \in Y_i$ and x_i is repaired. Since L provide coverage for the set $U = \{u_1, \cdots, u_n\}$, we can repair x_i for each $Y_i \in L$ to obtain a series that satisfies the residual constraint. Therefore, $\Delta(x, x') = card(L) \le \tau$.

Next we must demonstrate that any repair satisfying $\Delta(x, x') \le \tau$ provides $card(L) \le \tau$. Assume that there exists $card(L) > \tau$ when $\Delta(x, x') \le \tau$. Note that if we only repair point x_i but $u_j \notin Y_i$, then $r_j = n + 1$ or $n + 1 + x_{i-1} > n$. Therefore, we need to repair x_i where $u_j \in Y_i$ to make r_j satisfy the residual constraint. So $\Delta(x, x') \ge card(L)$ is a necessary condition if the entire series to be repaired, i.e., $\Delta(x, x') \ge card(L) > \tau$, which obviously contradict the initial assumption. In conclusion, we have demonstrated the NP-completeness of the decision problem at hand.

1.2 Proposition 3.1

Proof. First, consider the trend component t_i = median$\left(\left[x_{i+j};\ \lfloor -\frac{m-1}{2} \rfloor \le j \le \lfloor \frac{m-1}{2} \rfloor\right]\right)$. We define $A_k = \left[x_{k+l};\ \lfloor -\frac{m-1}{2} \rfloor \le l \le \lfloor \frac{m-1}{2} \rfloor\right]$ as the trend estimation vector of the original series with error x_i, where x_k is the k-th element of the series. Similarly, we define $B_k = \left[x_{k+l};\ \lfloor -\frac{m-1}{2} \rfloor \le l \le \lfloor \frac{m-1}{2} \rfloor \wedge l \neq i, x_i^*\right]$ as

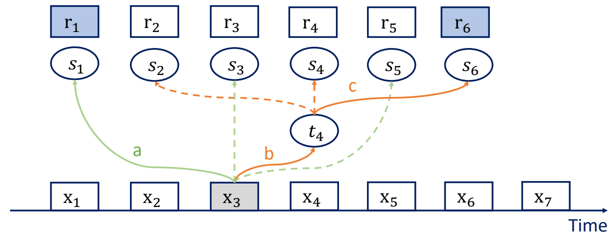

Figure 1: Transformation for the example set $Y_3 = \{u_1, u_6\}$

the trend estimation vector of the clean series. Since $x_i^* \ge \max(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$, x_i^* is the largest number in B_k. Further, since $x_i > x_i^*$, x_i is also the largest number in A_k. Considering that the median is the value separating the higher half from the lower half of the data, and the order of the elements in A_k and B_k is the same, the median median A_k and median B_k is also the same, i.e., the trend component on $[x_1, \cdots, x_i, \cdots, x_j, \cdots, x_n]$ with errors is exactly the same as that on the clean series $[x_1, \cdots, x_i^*, \cdots, x_j^*, \cdots, x_n]$.

Then consider the seasonal component s_i = median$\left(\left[d_j;\ 1 \le j \le n \wedge (j-i) \mid m\right]\right)$. We define $C_k = \left[d_l;\ 1 \le l \le n \wedge (l-k) \mid m\right]$ as the trend estimation vector of the original series with error x_i, and $B_k = \left[d_l;\ 1 \le j \le n \wedge (l-k) \mid m \wedge l \neq i, x_i^*\right]$ as the trend estimation vector of the clean series. Similarly to the trend component, since $x_i^* \ge \max(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$, x_i^* is the largest number in D_k. Further, since $x_i > x_i^*$, x_i is also the largest number in C_k. Therefore, the order of the elements in C_k and D_k is the same. Then the median median C_k and median D_k is also the same, i.e., the seasonal component on $[x_1, \cdots, x_i, \cdots, x_j, \cdots, x_n]$ with errors is exactly the same as that on the clean series $[x_1, \cdots, x_i^*, \cdots, x_j^*, \cdots, x_n]$. Since the trend term and the seasonal term of the clean sequence and the error sequence are equal, considering that $r_i = x_i - t_i - s_i$, the residual component is also equal.

For the minimum value x_j, there is a similar conclusion. In summary, the error-tolerant decomposition by Formulas 5 and 6 on $[x_1, \cdots, x_i, \cdots, x_j, \cdots, x_n]$ with errors is exactly the same as that on the clean series $[x_1, \cdots, x_i^*, \cdots, x_j^*, \cdots, x_n]$. □

1.3 Proposition 3.3

Proof. Let $W_k = \{x_l;\ ((l-k) \mid m) \wedge (l \neq k)\} \cup \{x_l;\ 1 \le l \le m\}$ for $1 \le k \le m$ in a original series, and $Y_k = W_k - \{x_k\} + \{x_k^*\}$ for $1 \le k \le m$ in a clean series. Since component extension $t_k = \text{median}(\{x_l; 1 - \lfloor -\frac{m-1}{2} \rfloor \le l \le k + \lfloor \frac{m-1}{2} \rfloor\} \cup \{t_{1+\lfloor \frac{m-1}{2} \rfloor} + s_l; m + k - \lfloor -\frac{m-1}{2} \rfloor \le l < m + 1 - \lfloor -\frac{m-1}{2} \rfloor\})$, t_k is obtained by calculating the median of the elements in W_k multiple times. Since $x_i^* \ge \max(W_k)$, x_i^* is the largest number in Y_k. Further, since $x_i > x_i^*$, x_i is also the largest number in W_k. Considering that the median is the value separating the higher half from the lower half of the data, and the order of the elements in W_k and Y_k is the same, the median in W_k and Y_k is also the same, i.e., the trend extension on $[x_1, \cdots, x_i, \cdots, x_j, \cdots, x_n]$ with errors is exactly the same as that on the clean series $[x_1, \cdots, x_i^*, \cdots, x_j^*, \cdots, x_n]$. For the minimum value x_j, there is a similar conclusion. □

1.4 Proposition 3.6

Proof. If $x_i^* \geq \max(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ is the only error and $x_i > x_i^*$, by Proposition 3.1, x_i can not affect the result of error tolerant decomposition. Furthermore, the repair value of x_i, i.e., x_i' is calculated from x_k which satisfies the residual constraint, so x_i' also satisfies the residual constraint. Therefore, the series after repairing x_i (1) has the same result of error tolerant decomposition, and (2) satisfies the residual constraint. As the result, the seasonal repair x' generated by Formulas 9 and 10 is always the optimal repair. For the minimum value x_j, there is a similar conclusion. In summary, for any errors $x_i > x_i^*$ and $x_j < x_j^*$ occurring on these two values, the seasonal repair x' generated by Formulas 9 and 10 is always the optimal repair with cost $\Delta(x, x') = 2$. □

1.5 Proposition 3.9

Proof sketch of Proposition 3.9. If only one cycle violates the residual constraint, then the components of the remaining data points with the same phase in the other cycles all satisfy the residual constraint. The seasonal repair generation takes the median of these residual components that satisfy the residual constraint, adds this to the seasonal and trend components which have the least influence from errors. This intuitively produces a repair that satisfies the residual constraint. □